

Schrödinger Equation As a General Optimization Algorithm

Xiaofei Huang

eGain Communications
Mountain View, CA 94043, U.S.A.
huangxiaofei@ieee.org

Abstract. One of the greatest scientific achievements of physics in the 20th century is the discovery of quantum mechanics. The Schrödinger equation is the most fundamental equation in quantum mechanics describing the time-based evolution of the quantum state of a physical system. It has been found that the time-independent version of the equation can be derived from a general optimization algorithm. Instead of arguing for a new interpretation and possible deeper principle for quantum mechanics, this paper elaborates a few points of the equation as a general global optimization algorithm. Benchmarked against randomly generated hard optimization problems, this paper shows that the algorithm significantly outperformed a classic local optimization algorithm. The former found a solution in one second with a single trial better than the best one found by the latter around one hour after one hundred thousand trials.

1 Introduction

Optimization is a core problem both in mathematics and computer science. It is a very active research area with many international conferences every year, a large amount of literature, and many researchers and practitioners across many fields for a wide range of applications. Combinatorial optimization [1,2] is a branch of optimization where the set of feasible solutions of problems is discrete, countable, and of a finite size. The general methods for combinatorial optimization are 1) local search [2,3], 2) simulated annealing [4,5], 3) genetic algorithms [6,7,8], 4) ant colony optimization [9], 5) tabu search [10], 6) branch-and-bound [11,12], and 7) dynamic programming [12]. The successful applications of different combinatorial optimization methods have been reported in solving a large variety of optimization problems in practice.

Cooperative optimization is a newly proposed general global optimization method [13,14,15] inspired by cooperation principles in team playing. Often times, individuals working together as a team can solve hard problems beyond the capability of any individual in the team. In its normal form, cooperative optimization has one and only one equilibrium and converges to it with an exponential rate regardless of initial conditions. It is also capable of identifying global optimal solutions so that it can efficiently terminate its search process.

Recently, it has been found (preprint:<http://arxiv.org/abs/quant-ph/0605220>) that the time-independent Schrödinger equation can be derived from a general form of cooperative optimization in a continuous-time version for continuous variables. In particular, when the cooperation among the agents is controlled at a certain level, the cooperative optimization algorithm becomes the time-independent Schrödinger equation when it reaches an equilibrium (a stationary state). Instead of arguing for a new interpretation and possible deeper principle for quantum mechanics, this paper elaborates a few points of the equation as a general global optimization algorithm, such as the existence of an equilibrium and the time-based evolution.

In the HDTV research lab, at Tsinghua University, Beijing, China, we found a noticeable improvement of the algorithm over the belief propagation [16,17], the state-of-the-art algorithm, at decoding some modern channel codes (a NP-hard problem similar to the integer programming problem). At solving those real-world hard optimization problems, classic optimization methods, including simulated annealing and genetic algorithms, failed to deliver a satisfying performance. Instead of using the problems biased towards some particular applications, this paper conducts the comparison over randomly generated hard optimization problems to check the generality of the Schrödinger equation as an optimization algorithm.

2 Cooperative Optimization and Schrödinger Equation

The fundamental difference between cooperative optimization and many classical optimization methods is at the very core of optimization, i.e., the way of making decisions for assigning decision variables. Classic one often times make a precise decision at assigning a variable at every time-instance of optimization, such as $x = 3$ for a time instance t . Such an assignment is precise at the sense that x can only be of the value 3, not any other ones. In contrast, cooperative optimization makes a soft decision, represented by a probability-like function called an assignment function, such as $\Psi(x, t)$, at every time-instance t . It says that at the time instance t , the variable x can be of any value with the likelihood measured by the function value $\Psi(x, t)$. A variable value of a higher function value is more likely to be assigned to the variable x than anyone of a lower function value.

If the function $\Psi(x, t)$ at time t is peaked at a specific value, say $x = 3$, while the others equal to zero, then the soft decision falls back to the classic precise decision, e.g., assigning the value 3 to the variable x ($x = 3$). Hence, the soft decision making is a generalization of the classic precise decision making.

Let $E(x_1, x_2, \dots, x_n)$, simply $E(x)$, be a multivariate objective function of n variables. Assume that $E(x)$ can be decomposed into n sub-objective functions $E_i(x)$, one for each variable, such that those sub-objective functions satisfying

$$E_1(x) + E_2(x) + \dots + E_n(x) = E(x) ,$$

and/or the minimization of $E_i(x_i)$ with respect to x_i also leads to the minimization of $E(x)$ for any i .

In terms of a multi-agent system, let us assign $E_i(x)$ as the objective function for agent i , for $i = 1, 2, \dots, n$. There are n agents in the system in total. The objective of the system is to minimize $E(x)$ and the objective of each agent i is to minimize $E_i(x)$.

A simple form of cooperative optimization is defined as an iterative update of the assignment function of each agent as follows:

$$\Psi_i(x_i, t) = \sum_{\sim x_i} \left(e^{-E_i(x)/\hbar} \prod_{j \neq i} p_j(x_j, t-1) \right), \quad \text{for } i = 1, 2, \dots, n, \quad (1)$$

where $\sum_{\sim x_i}$ stands for the summation over all variables except x_i and \hbar is a constant of a small positive value. $p_i(x_i, t)$ is defined as

$$p_i(x_i, t) = (\Psi_i(x_i, t))^\alpha / \sum_{x_i} (\Psi_i(x_i, t))^\alpha, \quad (2)$$

where α is a parameter of a positive real value.

By the definition, $p_i(x_i, t)$ just likes a probability function satisfying

$$\sum_{x_i} p_i(x_i, t) = 1.$$

It is, therefore, called the assignment probability function. It defines the probability-like soft decision at assigning variable x_i at the time instance t .

The original assignment function $\Psi_i(x_i, t)$, is called the assignment state function. That is, the state of agent i at the time instance t is represented by its assignment state function $\Psi_i(x_i, t)$. From Eq. 2 we can see that the assignment probability function $p_i(x_i, t)$ is defined as the assignment state function $\Psi_i(x_i)$ to the power α with normalization. To show the relationship, the assignment probability function $p_i(x_i, t)$ is also expressed as $(\bar{\Psi}_i(x_i, t))^\alpha$ in the following discussions with the bar standing for the normalization.

With this notation, the iterative update function (1) can be rewritten as

$$\Psi_i(x_i, t) = \sum_{\sim x_i} \left(e^{-E_i(x)/\hbar} \prod_{j \neq i} (\bar{\Psi}_j(x_j, t-1))^\alpha \right), \quad \text{for } i = 1, 2, \dots, n. \quad (3)$$

Without loss of generality, let $e^{-E_i(x)/\hbar}$ be the utility function of agent i in terms of game theory. It is important to note that when the parameter α is of a sufficient value, i.e., $\alpha \rightarrow \infty$, the simple form (3) converges to an equilibrium if and only if it is also a Nash equilibrium (preprint: <http://arxiv.org/abs/0901.3615>). Nash Equilibrium is arguably the most important concept in game theory, critical for understanding a common scenario in game playing. It offers the mathematical foundation for social science and economy.

By substituting Eq. 1 into Eq. 2, we have a mapping from a set of assignment probability functions to itself. Because the set is compact and the mapping is

continuous, so a fixed point exists based on Brouwer fixed point theorem. Since a set of assignment state functions is uniquely defined by a set of assignment probability functions by Eq. 1, We can conclude that there exists at least one set of assignment state functions $\{\Psi_1^*(x_1), \Psi_2^*(x_2), \dots, \Psi_n^*(x_n)\}$ such that

$$\Psi_i^*(x_i) = \sum_{\sim x_i} \left(e^{-E_i(x)/\hbar} \prod_{j \neq i} (\bar{\Psi}_j^*(x_j))^\alpha \right), \quad \text{for } i = 1, 2, \dots, n.$$

In particular, when $\alpha = 2$, the simple general form of cooperative optimization over the real domain \mathbb{R} given in (3) can be generalized further over the complex domain \mathbb{C} in a continuous-time version as follows

$$\hbar \frac{\partial \psi_i(x_i, t)}{\partial t} = -\frac{1}{Z_i(t)} \psi_i(x_i, t) \sum_{\sim x_i} \left(E_i(x) \prod_{j \neq i} |\psi_j(x_j, t)|^2 \right), \quad (4)$$

where $Z_i(t)$ is a normalization factor such that

$$\sum_{x_i} |\psi_i(x_i, t)|^2 = 1,$$

and $|\psi_i(x_i, t)|^2$ is defined as

$$|\psi_i(x_i, t)|^2 = \psi_i^*(x_i, t) \psi_i(x_i, t).$$

Following the notation from physics, denote $\psi_i(x_i, t)$ and a vector $|\psi_i(t)\rangle$. The equation (4) can be generalized further as follows

$$\hbar \frac{d}{dt} |\psi_i(t)\rangle = -\frac{1}{Z_i(t)} H_i |\psi_i(t)\rangle. \quad (5)$$

In the equation, H_i is a hermitian matrix which defines the local energy (objective) function for agent i in a more general form. In particular, for the case of (4), the hermitian matrix H_i reduces to a diagonal matrix with the diagonal elements as

$$\sum_{X_i \setminus x_i} \left(E_i(x) \prod_{j \neq i} |\psi_j(x_j, t)|^2 \right), \quad \text{for every value of } x_i.$$

The function $\psi_i(x_i, t)$ is also called a wavefunction in physics. It is important to note that the equation (5) is the dual equation of the Schrödinger equation:

$$i\hbar \frac{d}{dt} |\psi_i(t)\rangle = H_i |\psi_i(t)\rangle,$$

where -1 is replaced by the imaginary unit $i (= \sqrt{-1})$ and the normalization factor $Z_i(t)$ is not required since the equation is unitary, which means that the total norm of the wavefunction $\psi_i(x_i, t)$ is reserved, i.e.,

$$\sum_{x_i} |\psi_i(x_i, t)|^2 = \text{const}, \quad \text{for any } t.$$

When the dynamic equation (5) reaches a stationary point (equilibrium), it has been found out that the equation becomes the time-independent Schrödinger equation, one of the most important equations in quantum mechanics. That is

$$e_i | \psi_i(x_i, t) \rangle = H_i | \psi_i(x_i, t) \rangle ,$$

where e_i can only be any one of the eigenvalues of H_i .

If the local energy function H_i is time-independent, then the time-based evolution of the state function $| \psi_i(t) \rangle$ is

$$| \psi_i(t) \rangle = \frac{1}{Z_i(t)} e^{-\frac{1}{\hbar} H_i t} | \psi_i(0) \rangle .$$

Assume that $| \phi_1 \rangle, | \phi_2 \rangle, \dots, | \phi_N \rangle$ are the N eigenvectors of the Hamiltonian matrix H_i with corresponding eigenvalues as e_1, e_2, \dots, e_n , respectively. Then any initial state $| \psi_i(0) \rangle$ can be represented as the linear superposition of the N eigenvectors as follows

$$| \psi_i(0) \rangle = c_1 | \phi_1 \rangle + c_2 | \phi_2 \rangle + \dots + c_N | \phi_N \rangle ,$$

where c_n ($1 \leq n \leq N$ and $c_n \in \mathbb{C}$) are coefficients. The state function $| \psi_i(x_i, t) \rangle$ at time t is governed by

$$| \psi_i(t) \rangle = \frac{1}{Z_i(t)} \sum_{j=1}^N c_j e^{-\frac{1}{\hbar} e_j t} | \phi_j \rangle .$$

Benefit from almost a century's mathematical development driven quantum physics, we can extend the concept of a hermitian matrix to that of a hermitian operator to deal with the case when the variable x_i is continuous. In this case, the concept of an eigenvector will be extended to that of an eigenfunction.

Putting everything together, the pseudo-code of a quantum optimization algorithm is shown in Figure 1.

3 Experimental Comparisons

3.1 Constraint Optimization Problems

A large class of optimization problems has an objective function of the following form,

$$E(x) = \sum_{i=1}^n f_i(x_i) + \sum_{(i,j) \in \mathcal{N}} f_{ij}(x_i, x_j) . \quad (6)$$

The function $f_i(x_i)$ is called an unary function and the function $f_{ij}(x_i, x_j)$ is called a binary function. To note the collection of all defined binary functions, the set \mathcal{N} is used which contains non-ordered pairs of variable indices where each pair (i, j) corresponds to a defined binary function $f_{ij}(x_i, x_j)$.

```

1 begin
2   for every  $i$ , initialize  $\Psi_i(x_i)$  with random non-negative values and normalize it;
3   for  $k := 1$  to maximum_iteration_number do
4     begin
5       for  $i := 1$  to  $n$  do
6         begin
7           for each value of  $x_i$  do
8              $\Psi_i(x_i) := \sum_{\sim x_i} \left( e^{-E_i(x)/\hbar} \prod_{j \neq i} \bar{\Psi}_j^2(x_j) \right)$ ;
9             Let  $\tilde{x}_i := \arg \max_{x_i} \Psi_i(x_i)$ ;
10          end;
11          normalize  $\Psi_i(x_i)$  such that  $\sum_{x_i} |\psi_i(x_i)|^2 = 1$ ;
12        end;
13      return  $(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$  as the final solution;
14 end;

```

Fig. 1. A quantum optimization algorithm

Minimizing the objective function defined in (6) is referred to as a constraint optimization problem (binary COP) in AI. The unary function $f_i(x)$ is called an unary constraint on variable x_i and the binary function $f_{ij}(x_i, x_j)$ is called a binary constraint on variables x_i, x_j .

The constraint optimization problem is a very general formulation for many optimization problems arose from widely different fields. Examples are the famous traveling salesman problem, the weighted maximum satisfiability problem, the quadratic variable assignment problem, stereo matching in computer vision, image segmentation in image processing, and many more. The constraint optimization problem belongs to the class of NP-complete problems.

An objective function in form of (6) can be represented with an undirected graph. In the graph, each variable x_i is represented by a node and each binary constraint $f_{i,j}(x_i, x_j)$ is represented by an undirected edge connecting variable node x_i to variable node x_j . In this graphic representation, the degree of a node is measured by the total number of edges connected to it.

Each variable node x_i can have a number of neighboring variable nodes. Let $\mathcal{N}(i)$ be the set of the indices of the neighboring variables of x_i . By definition,

$$\mathcal{N}(i) = \{j | (i, j) \in \mathcal{N}\} .$$

Using the notations, we can define a local objective function for agent i as

$$E_i(x) = f_i(x_i) + \sum_{j \in \mathcal{N}(i)} f_{ij}(x_i, x_j), \quad \text{for } i = 1, 2, \dots, n . \quad (7)$$

Obviously, minimizing the local objective function $E_i(x)$ with respect to x_i also leads to the minimization of the global objective function $E(x)$.

3.2 A Classic Local Optimization Algorithm

Local search is one of the classic methods for attacking hard combinatorial optimization problems. It also plays a fundamental role at understanding many other optimization methods, such as simulated annealing, genetic algorithms, and tabu search. Started from a randomly generated initial solution, local search iteratively tries to replace the current solution by a better one in the neighborhood of the current solution until no further improvement is possible.

The improvement of the local search algorithm choosing in the experiments is achieved by adjusting the value of each variable while other variables have their values fixed. Specifically, assume that the current solution is \tilde{x} and x_i is the variable with its value to be adjusted. Its value in the current solution is \tilde{x}_i . Let x_{-i} denote all variables except x_i . In this case, the neighborhood is defined as (x_i, \tilde{x}_{-i}) where $x_i \in D_i$, the domain of x_i . The adjustment is achieved by finding the best value for x_i , denote as x_i^* , such that

$$E(x_i^*) = \min_{x_i} E(x_i, \tilde{x}_{-i}) ,$$

If $x_i^* \neq \tilde{x}_i$, then replacing \tilde{x}_i by x_i^* . Otherwise, keeping \tilde{x}_i as the value of x_i . If $x_i^* = \tilde{x}_i$ and it is true for every variable x_i , then no solution improvement is possible by adjusting any variable. In this case, the local search algorithm is called reaching a local optimal solution since it is optimal with respect only to its neighbors. It terminates the search process and returns the current solution \tilde{x} as the final one.

In particular, if $E(x)$ can be decomposed into n sub-objective functions $E_i(x)$ and minimizing $E_i(x)$ with respect to x_i also leads to minimizing $E(x)$, then the above optimization problem becomes

$$E_i(x_i^*) = \min_{x_i} E_i(x_i, \tilde{x}_{-i}) . \quad (8)$$

A pseudo-code of the above local search algorithm is given in Figure 2.

The multi-restart local search algorithm simply calls the above local search algorithm multiple times and picks the best solution (the best local optimal solution) as the final solution. In other words, it discovers multiple local optimal solutions and keeps the best one.

3.3 The Experimental Results

The size of a randomly generated instance is controlled by the total number of variables and the total number of values for each variable. The edges connecting to each variable node are randomly selected. If each node is connected to all other nodes, then the graph is a full graph. Otherwise, if it is connected to a small percentage of other nodes, then the graph is sparsely connected. Therefore, we can use the average node degree to control the graphical structure of each generated instance.

```

1  begin
5    randomly generate a solution  $\tilde{x}$ ;
6    repeat
7       $local\_minimum\_flag := true$ ;
8      for  $i := 1$  to  $n$  do
9        begin
10         find  $x_i^*$  such that  $E(x_i^*) = \min_{x_i} E(x_i, \tilde{x}_{-i})$ ;
11         if  $x_i^* \neq \tilde{x}_i$  then
12           begin
13              $\tilde{x}_i := x_i^*$ ;
14              $local\_minimum\_flag := false$ ;
15           end;
16         end;
17       until  $local\_minimum\_flag = true$ ;
20     return  $\tilde{x}$  as the final solution;
21 end;

```

Fig. 2. The local search algorithm

The function values of the constraints $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ are uniformly sampled from the interval $[0, 1]$. Because of the randomness, it is hard to apply any domain-specific heuristics for the advantage of any optimization algorithm.

In the first set of experiments, ten instances of the constraint optimization problem are generated. Each instance has 121 variables, 50 values for each variable, and the average node degree is six. Each instance has an enormous number of feasible solutions, bigger than the total number of atoms in the universe, posing a challenging optimization problem.

For each of those ten randomly generated instances, the best solution of the multi-restart local search algorithm (MRLS) after 100,000 trials is compared with the solution found by the quantum optimization algorithm (QOA) (Fig. 1) with a single trial. The parameters of the algorithm are set as $\hbar = 1$ and $maximum_iteration_number = 20$. The results are given in the following table. From it we can see that the quantum optimization algorithm yields better results within a much shorter time than the multi-restart local search algorithm. In the table, *variable#* represents the total number of variables and *value#* the total number of values for each variable.

MRLS with 100,000 trials <i>vs</i> QOA with a single trial variable#=121, value#=50, average_node_degree=6				
Instance#	MRLS:cost (time)	QOA:cost (time)	improvement	
1	153.11 (3001 sec)	144.77 (1.061 sec)	5.76%	
2	153.92 (3067 sec)	144.02 (1.046 sec)	6.87%	
3	152.12 (2994 sec)	136.84 (1.030 sec)	11.17%	
4	153.53 (3072 sec)	144.34 (1.045 sec)	6.37%	
5	154.95 (3013 sec)	145.22 (1.061 sec)	6.70%	
6	148.17 (3012 sec)	140.37 (1.029 sec)	5.55%	
7	147.90 (2905 sec)	138.83 (1.014 sec)	6.54%	
8	171.03 (3275 sec)	158.82 (1.138 sec)	7.69%	
9	154.70 (3013 sec)	145.59 (1.061 sec)	6.26%	
10	145.14 (2972 sec)	130.19 (1.030 sec)	11.49%	

The following table lists the experimental results with another set of instances where their sizes are different from the previous ones. The total number of variables is increased to 1001 and the total number of values for each variable is reduced to 10. From the table we can see that the quantum optimization algorithm still outperforms the multi-restart local search algorithm both in quality and speed.

MRLS with 10,000 trials <i>vs</i> QOA with a single trial variable#=1001,value#=10,average_node_degree=10				
Instance#	MRLS:cost (time)	QOA:cost (time)	improvement	
1	3269.97 (41.6 sec)	3102.63 (1.373 sec)	5.39%	
2	3221.61 (4126 sec)	3084.81 (1.415 sec)	4.43%	
3	3237.84 (4179 sec)	3090.48 (1.389 sec)	4.77%	
4	3270.37 (4134 sec)	3159.82 (1.435 sec)	3.50%	
5	3267.66 (4113 sec)	3109.14 (1.373 sec)	5.10%	
6	3307.75 (4108 sec)	3204.13 (1.404 sec)	3.23%	
7	3248.23 (4073 sec)	3153.07 (1.389 sec)	3.02%	
8	3273.33 (4077 sec)	3146.69 (1.388 sec)	4.02%	
9	3300.05 (4117 sec)	3188.34 (1.388 sec)	3.50%	
10	3269.44 (4126 sec)	3141.70 (1.454 sec)	4.07%	

4 Conclusions

The Schrödinger equation can be converted into a general purpose optimization algorithm. The original equation is in a continuous-time version while the one presented in this paper is in a discrete-time version so that it can be easily implemented by computer software or hardware. Benchmarked against randomly generated hard optimization problems, the quantum optimization algorithm significantly outperformed a classic multi-restart local search algorithm by several orders of magnitude. Together with many other existing optimization methods, the quantum optimization can be served as a new method to attack hard optimization problems.

References

1. Papadimitriou, C.H., Steiglitz, K., eds.: Combinatorial Optimization. Dover Publications, Inc. (1998)
2. Pardalos, P., Resende, M.: Handbook of Applied Optimization. Oxford University Press, Inc. (2002)
3. Michalewicz, Z., Fogel, D.: How to Solve It: Modern Heuristics. Springer-Verlag, New York (2002)
4. Kirkpatrick, Gelatt, C., Vecchi, M.: Optimization by simulated annealing. *Science* **220** (1983) 671–680
5. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6** (1984) 721–741
6. Fogel, L., Owens, A., Walsh, M.: Artificial Intelligence through Simulated Evolution. John Wiley, New York (1966)
7. Hinton, G., Sejnowski, T., Ackley, D.: Genetic algorithms. *Cognitive Science* (1992) 66–72
8. Schwefel, H.P.: Evolution and Optimum Seeking. John Wiley and Sons, Inc. (1995)
9. Dorigo, M., Stützle, T.: Ant Colony Optimization. The MIT Press, Cambridge, Massachusetts, London, England (2004)
10. Golver, F., Laguna, M.: Tabu Search. Kluwer Academic Publishers (1997)
11. Lawler, E.L., Wood, D.E.: Branch-and-bound methods: A survey. *OR* **14** (1966) 699–719
12. Jr., E.G.C., ed.: Computer and Job-Shop Scheduling. Wiley-Interscience, New York (1976)
13. Huang, X.: A general framework for constructing cooperative global optimization algorithms. In: *Frontiers in Global Optimization. Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers (2004) 179–221
14. Huang, X.: Cooperative optimization for solving large scale combinatorial problems. In: *Theory and Algorithms for Cooperative Systems. Series on Computers and Operations Research*. World Scientific (2004) 117–156
15. Huang, X.: Cooperative optimization for energy minimization in computer vision: A case study of stereo matching. In: *Pattern Recognition, 26th DAGM Symposium*, Springer-Verlag, LNCS 3175 (2004) 302–309
16. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann (1988)
17. Kschischang, F.R., Frey, B.J., andrea Loeliger, H.: Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory* **47** (2001) 498–519
18. Luce, R.D., Raiffa, H.: Games and Decisions: Introduction and Critical Survey. Dover (1985)
19. Messiah, A.: Quantum Mechanics. Dover Publications, Inc., Mineola, New York (1999)
20. Tegmark, M., Wheeler, J.A.: 100 years of quantum mysteries. *Scientific American* (2001) 68–75
21. Seife, C.: Do deeper principles underlie quantum uncertainty and nonlocality? *Science* **309** (2005) 98
22. Daskalakis, C., Papadimitriou, C.H.: Three-player games are hard. In: *Electronic Colloquium on Computational Complexity (ECCC)*. (2005) TR05–139
23. Daskalakis, C., Goldberg, P., Papadimitriou, C.: The complexity of computing a nash equilibrium. In: *Electronic Colloquium on Computational Complexity (ECCC)*. (2005) TR05–115